

Overview of the Face Recognition Grand Challenge*

P. Jonathon Phillips¹, Patrick J. Flynn², Todd Scruggs³, Kevin W. Bowyer²
Jin Chang², Kevin Hoffman³, Joe Marques⁴, Jaesik Min², William Worek³

¹National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899

²Computer Science & Engineering Depart., U. of Notre Dame, Notre Dame, IN 46556

³SAIC, 4001 N. Fairfax Dr., Arlington, VA 22203

⁴The Mitre Corporation, 7515 Colshire Drive, McLean, VA 22102

Abstract

Over the last couple of years, face recognition researchers have been developing new techniques. These developments are being fueled by advances in computer vision techniques, computer design, sensor design, and interest in fielding face recognition systems. Such advances hold the promise of reducing the error rate in face recognition systems by an order of magnitude over Face Recognition Vendor Test (FRVT) 2002 results. The Face Recognition Grand Challenge (FRGC) is designed to achieve this performance goal by presenting to researchers a six-experiment challenge problem along with data corpus of 50,000 images. The data consists of 3D scans and high resolution still imagery taken under controlled and uncontrolled conditions. This paper describes the challenge problem, data corpus, and presents baseline performance and preliminary results on natural statistics of facial imagery.

1. Introduction

In the past few years, a number of new face recognition techniques have been proposed. The new techniques include recognition from three-dimensional (3D) scans, recognition from high resolution still images, recognition from multiple still images, multi-modal face recognition, multi-algorithm, and preprocessing algorithms to correct for illumination and pose variations. These techniques hold the potential to improve performance of automatic face recognition by an order of magnitude over FRVT 2002 [1].

The Face Recognition Grand Challenge (FRGC) is designed to achieve this increase in performance by pursuing development of algorithms for all of the above proposed methods.¹ Determining the merit of these techniques requires three components: sufficient data; a challenge prob-

lem that is capable of measuring an order of magnitude improvement in performance; and the infrastructure that supports an objective comparison among different approaches.

The FRGC addresses all three requirements. The FRGC data corpus consists of 50,000 recordings divided into training and validation partitions. The data corpus contains high resolution still images taken under controlled lighting conditions and with unstructured illumination, 3D scans, and contemporaneously collected still images.

The challenge problem ensures that researchers are working on sufficiently large problems and that results are comparable between different approaches. The FRGC challenge problem consists of six experiments. The experiments measure performance on still images taken with controlled lighting and background, uncontrolled lighting and background, 3D imagery, multi-still imagery, and between 3D and still images. The infrastructure ensures that results from different algorithms are computed on the same data sets and that performance scores are generated by the same protocol. To measure progress under FRGC, the Face Recognition Vendor Test (FRVT) 2005, an independent evaluation on sequestered data, will be conducted.

There is heated debate among face recognition researchers about which method or technique will have better performance. FRGC should provide answers to some of these questions. In fact, authors of this paper have opposing views on one key debate: Will recognition from 3D imagery be more effective than recognition from high resolution 2D imagery? In Section 7 we state five, sometimes conflicting, conjectures, and relate them to specific experiments that will allow for an assessment of the conjectures at the conclusion of the FRGC.

2. Design of Data Set and Challenge Problem

The design of the FRGC starts from performance in FRVT 2002, establishes a performance goal that is an order of

*Please direct correspondence to Jonathon Phillips, jonathon@nist.gov

¹This paper discusses ver2.0 of the FRGC. Ver1.0 was a small challenge problem designed to introduce researchers to the FRGC challenge problem protocol, procedures, and data formats.

magnitude greater, and then designs a data corpus and challenge problem that supports meeting the FRGC performance goal.

The starting point for measuring the increase in performance is the high computational intensity test (HCInt) of the FRVT 2002. The images in the HCInt corpus were taken indoors under controlled lighting. The performance point selected as the reference is a verification rate of 80% (error rate of 20%) at a false accept rate (FAR) of 0.1%. This is the performance level of the top three FRVT 2002 participants. An order of magnitude increase in performance is a verification rate of 98% (2% error rate) at the same fixed FAR of 0.1%.

A challenge to designing the FRGC is collecting sufficient data to measure an error rate of 2%. Verification performance is characterized by two statistics: verification and false accept rates. The false accept rate is computed from comparisons between faces of different people. These comparisons are called non-matches. In most experiments, there are sufficient non-match scores because the number of non-match scores is usually quadratic in the size of the data set. The verification rate is computed from comparisons between two facial images of the same person. These comparisons are called match scores. Because the number of match scores is linear in the data set size, generating a sufficient number of matches can be difficult.

For a verification rate of 98%, the expected verification error rate is one in every 50 match scores. To be able to perform advanced statistical analysis, 50,000 match scores are required. From 50,000 match scores, the expected number of verification errors is 1,000 (at the FRGC performance goal).

The challenge is to design a data collection protocol that yields 50,000 match scores. We accomplished this by collecting images for a medium number of people with a medium number of replicates. The proposed FRGC data collection called for collecting images of 200 subjects once a week for an academic year (see section 4 for specific numbers), which generates approximately 50,000 match scores.

The design, development, tuning and evaluation of face recognition algorithms requires three data partitions: training, validation, and testing. The FRGC challenge problem provides training and validation partitions to researchers. A separate testing partition is being collected and sequestered for an independent evaluation.

The representation, feature selection, and classifier training is conducted on the training partition. For example, in PCA-based and LDA-based face recognition, the subspace representation is learned from the training set. In support vector machine (SVM) based face recognition algorithms, the SVM classifier is trained on the data in the training partition.

Challenge problem experiments are constructed from

data in the validation partition. During algorithm development, repeated runs are made on the challenge problems. This allows researchers to assess the best approaches and tune their algorithms. Repeated runs produce algorithms that are tuned to the validation partition. An algorithm that is not designed properly will not generalize to another data set.

To obtain an objective measure of performance requires that results be computed on a separate test data set. The test partition measures how well an approach generalizes to another data set. By sequestering the data in test partition, participants cannot tune their algorithm or system to the test data. This allows for an unbiased assessment of algorithm and system performance.

The FRGC experimental protocol is based on the FERET and FRVT 2002 testing protocols. For an experiment, the input to an algorithm is two sets of images: target and query sets. Images in the target set represent facial images known to a system. Images in the query set represent unknown images presented to a system for recognition. The output from an algorithm is a similarity matrix, in which each element is a similarity score that measures the degree of similarity between two facial images. The similarity matrix is comprised of the similarity scores between all pairs of images in the target and query matrices. Verification scores are computed from the similarity matrix.

3. Description of Data Set

Data for the FRGC was collected at the University of Notre Dame. The FRGC data corpus is part of an ongoing multimodal biometric data collection.

A *subject session* is the set of all images of a person taken each time a person's biometric data is collected. For the FRGC data for a subject session consists of four controlled still images, two uncontrolled still images, and one three-dimensional image. Figure 1 shows a set of images for one subject session. The controlled images were taken in a studio setting, are full frontal facial images taken under two lighting conditions (two or three studio lights) and with two facial expressions (smiling and neutral). The uncontrolled images were taken in varying illumination conditions; e.g., hallways, atria, or outdoors. Each set of uncontrolled images contains two expressions, smiling and neutral. The 3D images were taken under controlled illumination conditions appropriate for the Vivid 900/910 sensor, not the same as the conditions for the controlled still images. In the FRGC, 3D images consist of both range and texture channels. The Vivid sensor acquires the texture channel just after the acquisition of the shape channel. This can result in subject motion that can cause poor registration between the texture and shape channels.

The still images were taken with a 4 Megapixel Canon

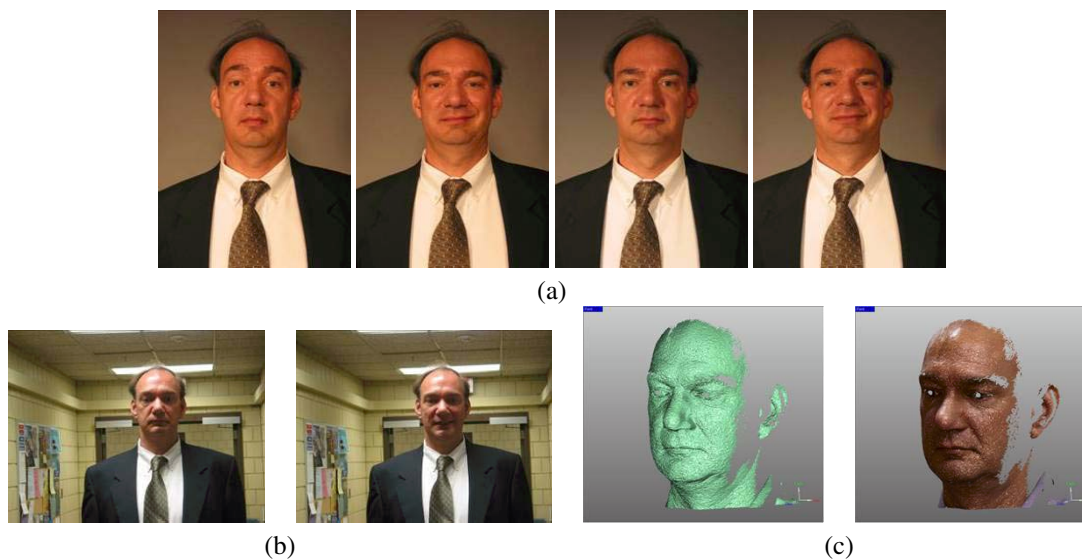


Figure 1: Images from one subject session. (a) Four controlled stills, (b) two uncontrolled stills, and (c) 3D shape channel and texture channel pasted on 3D shape channel.

Table 1: Size of faces in the validation set imagery broken out by category. Size is measured in pixels between the centers of the eyes. Reported is mean, median, and standard deviation.

	Mean	Median	Std. Dev.
Controlled	261	260	19
Uncontrolled	144	143	14
3D	160	162	15

PowerShot G2. Images are either 1704x2272 pixels or 1200x1600 pixels. Images are in JPEG format and storage sizes range from 1.2 Mbytes to 3.1 Mbytes².

The 3D images were acquired by a Minolta Vivid 900/910 series sensor. The Minolta Vivid 900/910 series is a structured light sensor that takes a 640 by 480 range sampling and a registered color image. Subjects stood or sat approximately 1.5 meters from the sensor. The images for the FRGC were not acquired using the 900/910's reduced resolution "fast mode", so that the images were acquired with resolution beyond that typically used in 3D face recognition today.

Table 1 summarizes the size of the faces for the uncontrolled, uncontrolled, and 3D image categories. For comparison, the average distance between the centers of the eyes in the FERET database is 68 pixels with a standard deviation of 8.7 pixels.

²The identification of any commercial product or trade name does not imply endorsement or recommendation by the National Institute of Standards and Technology, Notre Dame, SAIC, or Mitre.

The data for the FRGC experiments was divided into training and validation partitions. The data in the training partition was collected in the 2002-2003 academic year. From the training partition, two training sets were distributed. The first is the *large still training set*, which is designed for training still face recognition algorithms. The large still training set consists of 12,776 images from 222 subjects, with 6,388 controlled still images and 6,388 uncontrolled still images. The large still training set contains from 9 to 16 subject sessions per subject, with the mode being 16. The second training set is the *3D training set* that contains 3D scans, and controlled and uncontrolled still images from 943 subject sessions. The 3D training set is for training 3D and 3D to 2D algorithms. Still face recognition algorithms can be training from the 3D training set when experiments that compare 3D and still algorithms need to control for training. There are separate still and 3D training sets because 3D facial scans were only collected for part of the 2002-2003 academic year.

Images in the validation partition were collected during the 2003-2004 academic year. The validation set contains images from 466 subjects collected in 4,007 subject sessions. The demographics of the validation partition broken out by sex, age, and race are given in Figure 2. The validation partition contains from 1 to 22 subject sessions per subject (see Figure 3).

4. Description of Experiments

The experiments in FRGC ver2.0 are designed to advance face recognition in general with emphasis on 3D and high

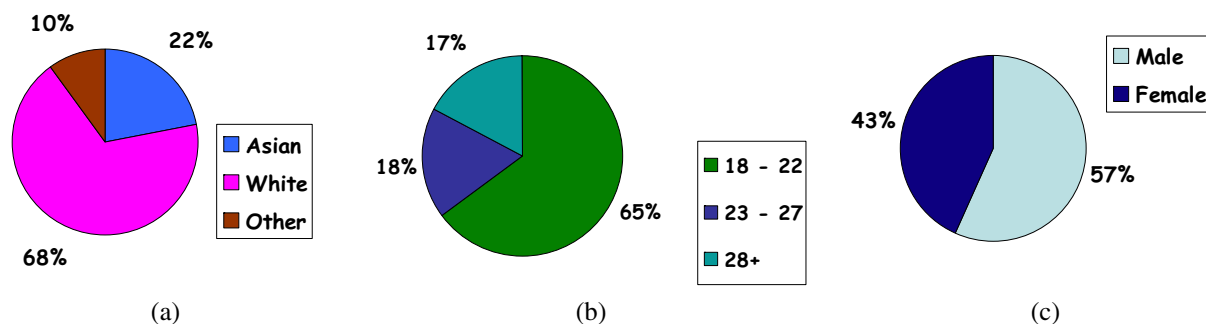


Figure 2: Demographics of FRGC ver2.0 validation partition by (a) race, (b) age, and (c) sex.

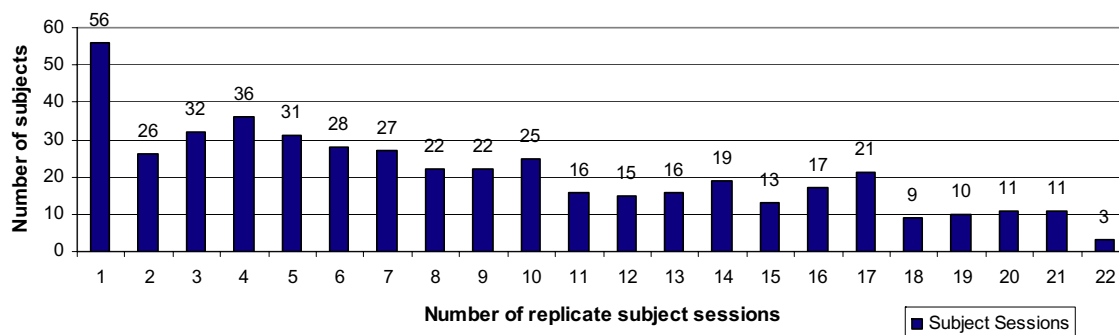


Figure 3: Histogram of the distribution of subjects for a given number of replicate subject sessions. The histogram is for the ver2.0 validation partition.

resolution still imagery. Ver2.0 consists of six experiments.

Experiment 1 measures performance on the classic face recognition problem: recognition from frontal facial images taken under controlled illumination. To encourage the development of high resolution recognition, all controlled still images are high resolution. In Experiment 1, the biometric samples in the target and query sets consist of a single controlled still image.

Recently, researchers have observed that multi-still images of a person can substantially improve performance [2][3]. Experiment 2 is designed to examine the effect of multiple still images on performance. In this experiment, each biometric sample consists of the four controlled images of a person taken in a subject session. The biometric samples in the target and query sets are composed of the four controlled images of each person from a subject session.

Recognizing faces under uncontrolled illumination has numerous applications and is one of the most difficult problems in face recognition. Experiment 4 is designed to measure progress on recognition from uncontrolled frontal still images. In Experiment 4, the target set consists of single controlled still images, and the query set consists of single

uncontrolled still images.

Proponents of 3D face recognition claim that 3D imagery is capable of achieving an order of magnitude increase in face recognition performance. Experiments 3, 5, and 6 examine different potential implementations of 3D face recognition. Experiment 3 measures performance when both the enrolled and query images are 3D. In Experiment 3, the target and query sets consist of 3D facial images.

One potential scenario for 3D face recognition is that the enrolled images are 3D and the target images are still 2D images. Experiment 5 explores this scenario when the query images are controlled while Experiment 6 examines the uncontrolled query image scenario. In both experiments, the target set consists of 3D images. In Experiment 5, the query set consists of a single controlled still. In Experiment 6, the query set consists of a single uncontrolled still. The size of each experiment in terms of training, target, and query set, and number of similarity scores is given in Table 2.

5. Baseline Performance

Baseline performance serves to demonstrate that a challenge problem can be executed, to provide a minimum level

Table 2: Size of ver2.0 experiments. For each experiment the size of the training, target, and query set is given. For the training set the number of biometric samples provided is divided into still and 3D images. The number of similarity scores in each experiment’s similarity matrix is provided.

Exp.	Training set size (still/3D)	Target set size	Query set size	No. sim scores (million)
1	12,776/-	16,028	16,028	257
2	12,776/-	4,007	4,007	16
3	-/943	4,007	4,007	16
4	12,776/-	16,028	8,014	128
5	3772/943	4,007	16,028	64
6	1886/943	4,007	8,014	32

of performance, and to provide a set of controls for detailed studies. A PCA-based face recognition was selected as the baseline algorithm [4].

The initial set of baseline performance results is given for Experiments 1, 2, 3, and 4. For Experiments 1, 2, and 4, baseline scores were computed from the same PCA-based implementation. In Experiment 2, a fusion module was added to handle multiple recordings in the biometric samples. Our implementation is based on Moon and Phillips [5] and Beveridge et al [6]. The algorithm was trained on a subset of 2,048 images from the large training set. The representation consists of the first 1,228 eigenfeatures (60% of the total eigenfeatures). See Section 6 for details on the selection of the training set size and number of eigenfeatures in the representation. All images were preprocessed by performing geometric normalization, masking, histogram equalization, and rescaling pixels to have mean zero and unit variance. All PCA spaces are whitened. The distance in nearest neighbor classifier is the cosine of the angle between two representations in a PCA-space. In Experiment 2, each biometric sample consists of four still images, and comparing two biometric samples involves two sets of four images. Matching all four images in both sets produces 16 similarity scores. For Experiment 2, the final similarity score between the two biometric samples is the average of the 16 similarity scores between the individual still images.

A set of baseline performance results is given for Experiment 3 (3D versus 3D face recognition). The baseline algorithm for the 3D scans consists of PCA performed on the shape and texture channels separately and then fused. Performance scores are given for each channel separately and for the shape and texture channels fused. We also fused the 3D shape channel and one of the controlled still images. The controlled still is taken from the same subject session as the 3D scan. Using the controlled still models a situation where superior still camera is incorporated into the 3D sensor. The baseline algorithm for the texture channel is same

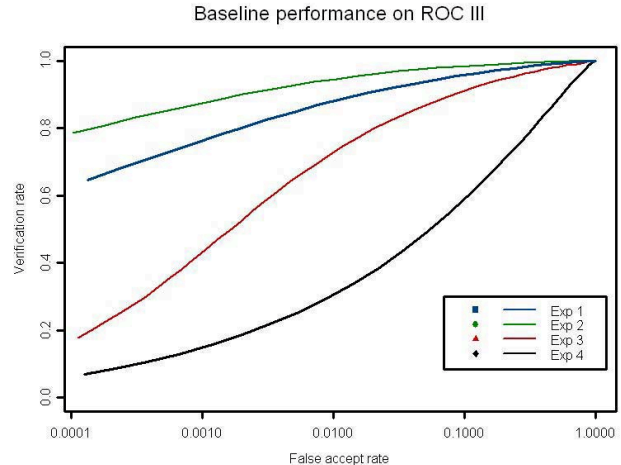


Figure 4: Baseline ROC performance for Experiments 1, 2, 3, and 4.

as in Experiment 1. The PCA algorithm adapted for 3D is based on Chang et al [2].

Baseline verification performance results for Experiments 1, 2, 3, and 4 are reported in Figure 4. Verification performance is computed from target images collected in the Fall semester and query images collected in the Spring semester. For these results, the time lapse between images is between two and ten months. Performance is reported on a receiver operator characteristic (ROC) that shows the trade-off between verification and false accept rates. The false accept rate axis is logarithmic. The results for Experiment 3 are based on fused shape and texture channels. The best baseline performance was achieved by multi-still images, followed by a single controlled still, then 3D scans. The most difficult category was the uncontrolled stills.

Figure 5 shows baseline performance for five configurations of the 3D baseline algorithms: fusion of 3D shape and one controlled still; controlled still; fusion of 3D shape and 3D texture; 3D shape; and 3D texture. The best result is achieved by fusing the 3D shape channel and one controlled still image. This result suggests that 3D sensors equipped with higher quality still cameras and illumination better optimized to still cameras may improve performance of 3D systems.

6. Facial Image Statistics

The vast majority of face recognition papers are concerned with presenting a novel algorithm and demonstrating that it has superior performance. However, few researchers investigate the fundamental properties of facial images or the underlying reasons why one class of algorithms is better than

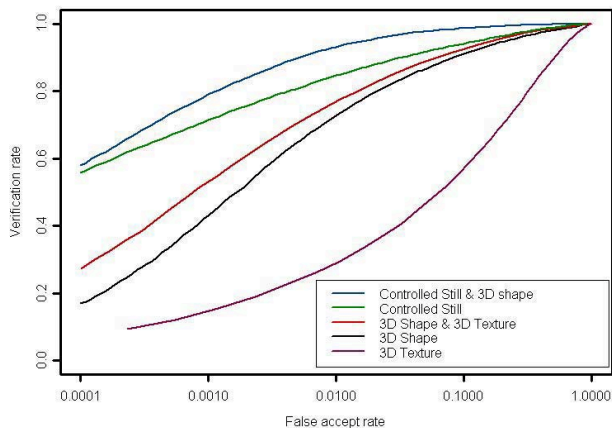


Figure 5: Baseline ROC performance for Experiment 3 component study. Performance is reported for fusion of 3D shape and one controlled still image; one controlled still image; fusion of 3D texture and 3D shape; 3D shape; and 3D texture.

another. A natural starting point for this is the statistics of facial images.

Similar to issues in image statistics of natural scenes, this section looks at facial images statistics and underlying properties of face processing [7]. Initial studies of facial image statistics could easily fill several papers. However, in this section, we will briefly introduce the concept by commenting on the effect of training set size on the eigenspectrum and performance.

Successful development of pattern recognition algorithms requires that one knows the distributional properties of objects being recognized. A natural starting point is PCA, which assumes the facial distribution has a multivariate Gaussian distribution in projection space. In keeping with Occam's razor, once the deficiencies of this simple model have been adequately identified, then it is time to examine more complex models. The deficiencies of PCA as a recognition algorithm have been documented. However, from a facial image statistics perspective, this model has not been examined.

In the first facial statistics experiment we examine the effect of the training set size on the eigenspectrum. If the eigenspectrum is stable, then variance of the facial statistics on the principal components is stable. The eigenspectrum was computed for five training sets of size 512, 1,024, 2,048, 4,096, and 8,192. All the training sets are subsets of the large still training set. The eigenspectra are plotted in Figure 6. The horizontal axis is the index for the eigenvalue on a logarithmic scale and the vertical axis is the eigenvalue on a

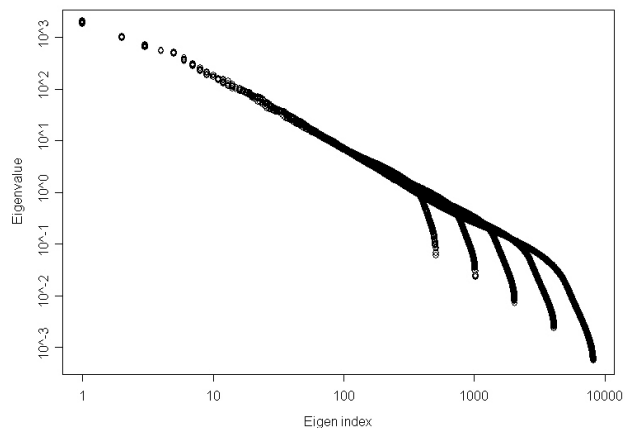


Figure 6: The spectrum of the eigenvalues for training sets of size 8,192, 4,096, 2,048, 1,024, and 512. For the low to mid order eigenvalues, the plots for all five spectra overlap. Each of the spectra have high order tails. The tails from left to right are from the 512, 1,024, 2,048, 4,096, 8,192 training sets.

logarithmic scale. The main part of the spectrum consists of the low to mid order eigenvalues. For all five eigenspectra, the main parts overlap.

The eigenvalues are estimates of the variance of the facespace distribution along the principal axes. Figure 6 shows that the estimates of the variances on the principal components are stable as the size of training set increases, excluding the tails. The main part of the eigenspectrum is approximately linear, which suggests that to a first order approximation there is a $1/f$ relationship between eigen-index and the eigenvalues.

The natural follow-up question to the eigenspectra observation is: Are the distribution of the eigen-coefficients similar across the training sets? We provide an initial answer to this question by plotting the estimated density for two coefficients (see Figure 7). The distributions for the first eigen-coefficients, Figure 7(a), are bimodal. None of the other coefficient distributions examined were bimodal. The distribution in Figure 7(b) is more representative of coefficient distributions. They have the same general distribution for each training set. These preliminary observations indicate that the distribution of coefficients deserves greater attention. The size and composition of the FRGC corpus makes studies of facial image statistics such as this one possible.

The effect of the size of the training set on performance in face recognition has not been studied. The largest standard training set in the literature is 501 images in the FERET Sep96 protocol. In the next experiment, we look at

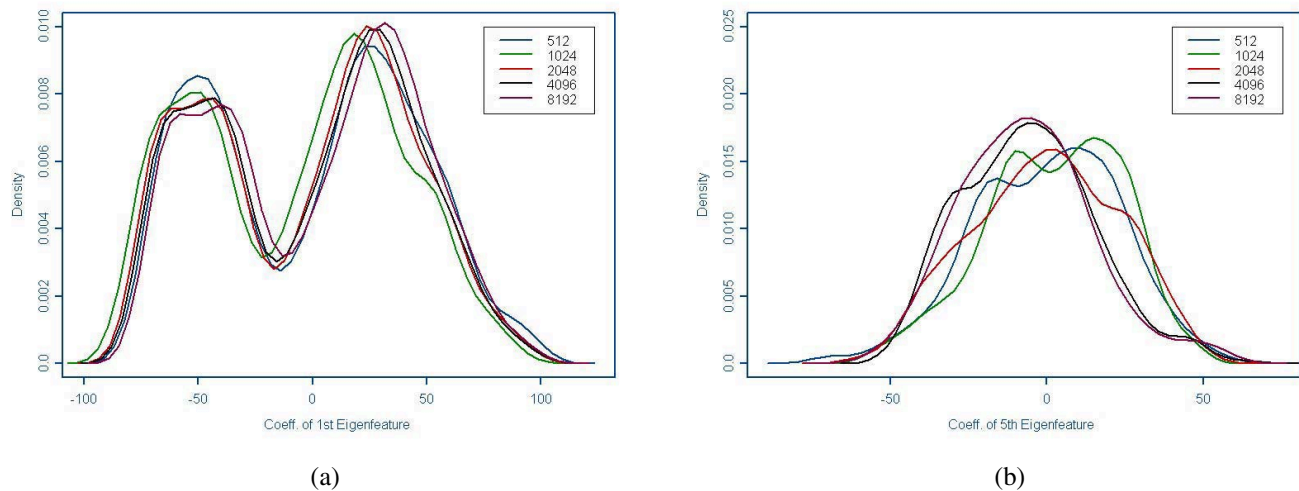


Figure 7: Estimated densities for the (a) 1st and (b) 5th eigen-coefficients for each training set (the numbers in the legend are the training set size). To generate the curve label 1024 in (a), a set of images were projected on the 1st eigenfeature generated the 1024 training set. The set of images projected onto the eigenfeatures was a subset of 512 image in common to all five training sets. All other curves were generated in a similar manner.

the effect of the size of the training set on PCA performance. Figure 8 reports performance on Experiment 1 for training sets of size 512, 1,024, 2,048, 4,096, and 8,192. Verification performance at a false accept rate of 0.1% is reported (vertical axis). The horizontal axis is the number of eigenfeatures in the representation. The eigenfeatures selected are the first n components. The training set of size 512 approximates the size of the training set in the FERET Sep96 protocol. This curve approximates what was observed in Moon and Phillips, where performance increases, peaks, and then decreases slightly [5]. Performance peaks for training sets of size 2,048 and 4,096 and then starts to decrease for the training set of size 8,192. For training sets of size 2,048 and 4,096, there is a large region where performance is stable. The training sets of size 2,048, 4,096, and 8,192 have tails where performance degrades to near zero.[6]

There are three immediate conclusions from this experiment: first, increasing the training set increases performance to a point; second, selection of the cutoff index is not critical; and third, performance of the training set of size 512 is a warning about drawing conclusions from too small a sample.

7. Conclusions and Conjectures

While the majority of face recognition researchers will agree that performance can be significantly increased, there is a contentious debate about how to achieve this goal. The biggest fault line is the divide between advocates for 3D face recognition and proponents of high resolution still im-

agery.

The FRGC challenge problems allow for these differences in opinion to be formulated as testable conjectures. In the design of the FRGC, we state five conjectures. The FRGC answers to these conjectures will not be final, but rather will be part of the process of building a scientific consensus. The answers to our conjectures will change over time and vary depending on the data set. The FRGC will be the first opportunity to test the validity of the conjectures. We relate Conjectures I to IV to the FRGC experiments³. Conjecture V characterizes the potential effects of the FRGC project on operational face recognition systems.

Conjectures I and II directly address the 3D versus 2D still debate. Conjectures III and IV address special cases of the 3D versus 2D still debate.

Conjecture I (Bowyer's)⁴: The shape channel of one 3D image is more powerful for face recognition than one 2D image.

One of the contentions in establishing the criteria for Conjecture I is specifying comparable images between the modes. In the spirit of compromise and scientific investigation (and at the risk of confusion in face recognition community) we list five criteria for Conjecture I.

Criterion I-A: Performance on Experiment 3 (shape

³Performance will be measured ROC III and differences in performance will be based on an appropriate statistical measure.

⁴The conjectures do not reflect the opinion of NIST, Notre Dame, or SAIC. The conjectures are designed to be a catalyst for establishing a scientific consensus.

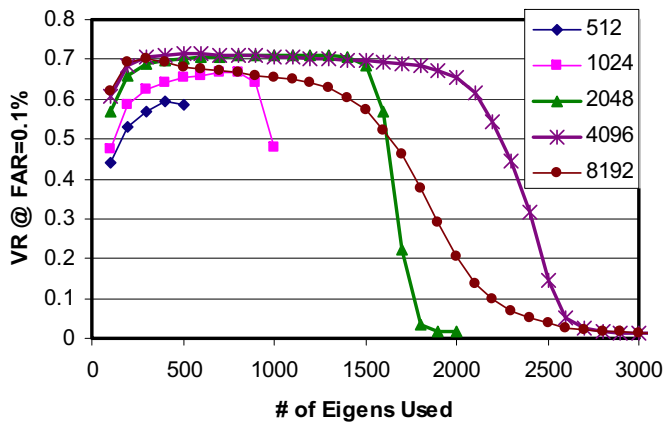


Figure 8: Verification performance as a function of the number of eigenfeatures in the representation for different size training sets. Verification performance at a false accept rate of 0.1% is reported. The numbers in the legend is the size of the training set. Performance is for the baseline PCA algorithm on experiment 1.

only) will be better than Experiment 3 (texture only).

Criterion I-B: Performance on Experiment 3 (shape only) will be better than Experiment 1 with the images scaled to 90 pixels between the centers of the eyes. The proposed ISO SC-37 facial image standard.

Criterion I-C: Performance on Experiment 3 (shape and texture) will be better than Experiment 1 with the images scaled to 90 pixels between the centers of the eyes.

Criterion I-D: Performance on Experiment 3 (shape only) will be better than Experiment 1 with the images at the original resolution.

Criterion I-E: Performance on Experiment 3 (shape and texture) will be better than Experiment 1 with the images at the original resolution.

Conjecture II (Phillips’): One high resolution 2D image is more powerful for face recognition than one 3D image.

Criteria: The opposite of criteria I-D and I-E.

Conjecture III: Using 4 or 5 well-chosen 2D face images is more powerful for recognition than one 3D face image or one multi-modal 3D+2D fusion.

Criterion III-A: Performance on Experiment 2 will be better than Experiment 3 (shape and texture).

Criterion III-B: Performance on Experiment 2 will be better than Experiment 3 (shape only).

Conjecture IV: The most promising aspect of 3D is addressing the case where the known images of a person are 3D biometric samples and the samples to be recognized are uncontrolled stills.

Criterion IV-A: Performance on Experiment 6 will be

better than Experiment 4.

Conjecture V: Solution to the FRGC will cause rethinking of how face recognition is deployed.

The recognition engine in a face recognition implementation is one part of a complex information technology system. Other key components of a face recognition implementation include image collection protocols; storage and transmissions infrastructure; and international standards for facial imagery. Current face recognition implementations are designed to work with a single still image of a person. The size of the face is 60 to 90 pixels between the centers of the eyes and image is compressed to approximately 10,000 bytes. Conjecture V posits a change in attitude in thinking about using face recognition. The results of FRGC may indicate that solutions to face recognition applications require 3D face scans, or multiple facial images, or high resolution images. Adapting to these conclusions has the potential to cause a rethinking and re-engineering of how face recognition is deployed in real-world applications.

Acknowledgments

The authors acknowledge and thank the Department of Homeland Security Science and Technology Directorate, Federal Bureau of Investigation, Intelligence Technology Innovation Center, and Technical Support Working Group for their support of the Face Recognition Grand Challenge.

References

- [1] P.J. Phillips, P.J. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, “Face recognition vendor test 2002: Evaluation report”, Tech. Rep. NISTIR 6965, National Institute of Standards and Technology, 2003, <http://www.frvt.org>.
- [2] Kyong I. Chang, Kevin W. Bowyer, and Patrick J. Flynn, “An evaluation of multi-modal 2d+3d face biometrics”, *IEEE Trans. PAMI*, vol. 27, no. 4, pp. 619–624, 2005.
- [3] P.J. Grother, “Face recognition vendor test 2002: Supplemental report”, Tech. Rep. NISTIR 7083, National Institute of Standards and Technology, 2004, <http://www.frvt.org>.
- [4] M. Turk and A. Pentland, “Eigenfaces for recognition”, *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [5] H. Moon and P. J. Phillips, “Computational and performance aspects of PCA-based face-recognition algorithms”, *Perception*, vol. 30, pp. 303–321, 2001.
- [6] J. R. Beveridge, D. Bolme, B. A. Draper, and M. Teixeira, “The CSU face identification evaluation system”, *Machine Vision and Applications*, 2004.
- [7] D. L. Ruderman, “The statistics of natural images”, *Network: Computation in Neural Systems*, vol. 5, pp. 517–548, 1994.